

Pengaruh Faktor-Faktor Kondisi Pada Suatu Negara Terhadap Angka Harapan Hidup

Kelompok 14

Laurensius Steven [1906352142]
Marko Chindranata [1906373885]
Nazhira Ghaisani [1906357805]
Samuel Tjahjono [1906375953]
Teresa Yubilea [1906357654]

Pendahuluan

Penelitian kami menggunakan data dari
<https://www.kaggle.com/kumarajarshi/life-expectancy-who>

Dataset tersebut berisikan *life expectancy* dan beberapa faktor kesehatan pada 183 negara dari tahun 2000 sampai 2015. Sumber data berasal dari data milik WHO, sedangkan data mengenai keadaan ekonomi diambil dari website PBB.

Terdapat beberapa *missing value*, di mana sebagian besarnya adalah pada fitur populasi, Hepatitis B, dan GDP. Hal ini terjadi pada beberapa negara yang kurang banyak dikenal, seperti Vanuatu, Tonga, Togo, Cabo Verde, dan lainnya

1. Rumusan Masalah Penelitian: Apa saja faktor-faktor yang mempengaruhi tingkat harapan hidup pada suatu negara?
2. Objektif model: Mengetahui faktor apa saja yang mempengaruhi tingkat harapan hidup pada suatu negara yang didasarkan pada beberapa faktor kondisi suatu negara, seperti status perkembangan sebuah negara, tingkat mortalitas per 1000 populasinya, dan lainnya.

Data Awal

Status	Life.expectancy	Adult.Mortality	infant.deaths	percentage.expenditure	Measles	under.five.deaths	Polio	Diphtheria	HIV.AIDS
Developing	65.0	263	62	71.27962	1154	83	6	65	0.1
Developing	77.8	74	0	364.97523	0	0	99	99	0.1
Developing	75.6	19	21	0.00000	63	24	95	95	0.1
Developing	52.4	335	66	0.00000	118	98	7	64	1.9
Developing	76.4	13	0	0.00000	0	0	86	99	0.2
Developing	76.3	116	8	0.00000	0	9	93	94	0.1
Developing	74.8	118	1	0.00000	33	1	96	94	0.1
Developed	82.8	59	1	0.00000	74	1	93	93	0.1
Developed	81.5	65	0	0.00000	309	0	93	93	0.1
Developing	72.7	118	5	0.00000	0	6	98	96	0.1

Struktur Data ->

```
'data.frame': 183 obs. of 10 variables:
 $ Status      : chr  "Developing" "Developing" "Developing" "Developing" ...
 $ Life.expectancy : num  65 77.8 75.6 52.4 76.4 76.3 74.8 82.8 81.5 72.7 ...
 $ Adult.Mortality : int   263 74 19 335 13 116 118 59 65 118 ...
 $ infant.deaths  : int    62 0 21 66 0 8 1 1 0 5 ...
 $ percentage.expenditure: num  71.3 365 0 0 0 ...
 $ Measles        : int  1154 0 63 118 0 0 33 74 309 0 ...
 $ under.five.deaths : int   83 0 24 98 0 9 1 1 0 6 ...
 $ Polio          : int    6 99 95 7 86 93 96 93 93 98 ...
 $ Diphtheria     : int   65 99 95 64 99 94 94 93 93 96 ...
 $ HIV.AIDS       : num  0.1 0.1 0.1 1.9 0.2 0.1 0.1 0.1 0.1 0.1 ...
```

Exploratory Data Analysis – Model Awal

```
> summary(mod1)

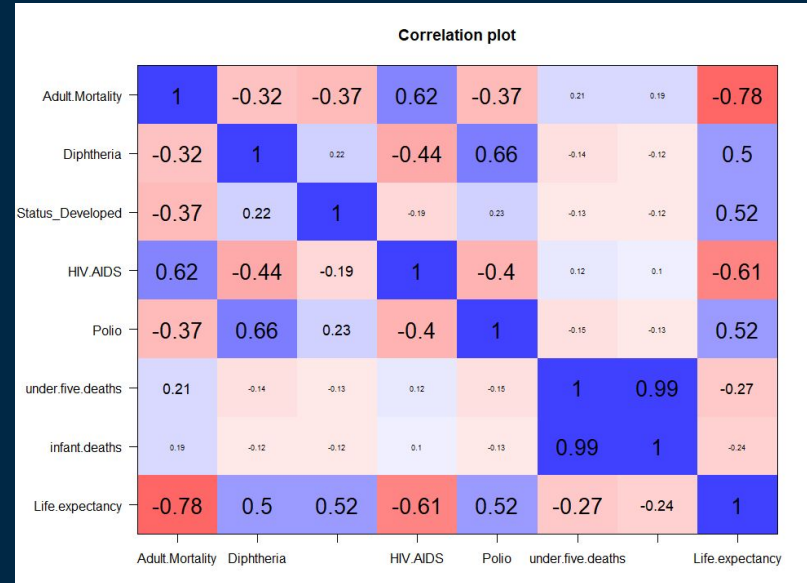
Call:
lm(formula = Life.expectancy ~ Adult.Mortality + Diphtheria +
    Status + HIV.AIDS + Polio + under.five.deaths + infant.deaths,
    data = df2)

Residuals:
    Min       1Q   Median       3Q      Max
-15.814  -2.329   0.447   2.225   9.653

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  75.604258   1.804547  41.897 < 2e-16 ***
Adult.Mortality -0.041119   0.004307  -9.547 < 2e-16 ***
Diphtheria    0.047859   0.019067   2.510  0.0130 *
StatusDeveloping -5.170456   0.871665  -5.932 1.57e-08 ***
HIV.AIDS     -0.716312   0.308881  -2.319  0.0215 *
Polio         0.038945   0.017218   2.262  0.0249 *
under.five.deaths -0.067973  0.027415  -2.479  0.0141 *
infant.deaths  0.079635   0.035126   2.267  0.0246 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.103 on 175 degrees of freedom
Multiple R-squared:  0.7547,    Adjusted R-squared:  0.7448
F-statistic: 76.9 on 7 and 175 DF, p-value: < 2.2e-16
```

Hasil model yang pertama kali dibuat berdasarkan pemilihan variabel *step-wise* AIC.



Tabel korelasi antar variabel dari model

Hasil VIF

Adult.Mortality	Diphtheria	Status	HIV.AIDS	Polio	under.five.deaths	infant.deaths
1.908520	1.930050	1.191438	1.864674	1.883685	94.365739	93.390019

Exploratory Data Analysis - Model 2

```
Call:
lm(formula = Life expectancy ~ ., data = df5)

Residuals:
    Min       1Q   Median       3Q      Max
-16.7711  -2.3318   0.4332   2.5149   9.5917

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  70.253188   1.672517  42.004 < 2e-16 ***
Adult.Mortality -0.042227   0.004346  -9.716 < 2e-16 ***
Diphtheria    0.050932   0.019303   2.639 0.00907 **
status_Developed 5.070042   0.883366   5.739 4.08e-08 ***
HIV.AIDS     -0.794686   0.311721  -2.549 0.01165 *
Polio        0.039809   0.017464   2.279 0.02384 *
infant.deaths -0.006968   0.003773  -1.847 0.06650 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.163 on 176 degrees of freedom
Multiple R-squared:  0.746,    Adjusted R-squared:  0.7374
F-statistic: 86.17 on 6 and 176 DF,  p-value: < 2.2e-16
```

Hasil model 2 setelah
menghilangkan variabel yang
berkorelasi tinggi

Adult.Mortality	Diphtheria	status_Developed	HIV.AIDS	Polio	infant.deaths
1.887975	1.921893	1.188866	1.845146	1.882914	1.047165

Hasil VIF

Exploratory Data Analysis – Uji Interaksi

```
Call:
lm(formula = Life.expectancy ~ Adult.Mortality + Diphtheria +
    Status_Developed + HIV.AIDS + Polio + infant.deaths + Adult.Mortality:HIV.AIDS +
    Adult.Mortality:Diphtheria + Diphtheria:Polio + Adult.Mortality:Polio,
    data = df5)

Residuals:
    Min       1Q   Median       3Q      Max
-15.7524  -2.1503   0.3342   2.2591   9.2688

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    81.7022745   6.3317762   12.904 < 2e-16 ***
Adult.Mortality -0.0526500   0.0175492   -3.000 0.003100 **
Diphtheria      -0.0040780   0.0595563   -0.068 0.945489
Status_Developed 4.2924500   0.8433606    5.090 9.33e-07 ***
HIV.AIDS        -5.2849497   1.2444517   -4.247 3.54e-05 ***
Polio           -0.1424675   0.0840335   -1.695 0.091816 .
infant.deaths   -0.0058872   0.0035222   -1.671 0.096449 .
Adult.Mortality:HIV.AIDS 0.0134301   0.0036473    3.682 0.000309 ***
Adult.Mortality:Diphtheria -0.0003513   0.0001486   -2.364 0.019209 *
Diphtheria:Polio 0.0014249   0.0006547    2.176 0.030882 *
Adult.Mortality:Polio 0.0003684   0.0002220    1.659 0.098878 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.861 on 172 degrees of freedom
Multiple R-squared:  0.7865,    Adjusted R-squared:  0.7741
F-statistic: 63.36 on 10 and 172 DF,  p-value: < 2.2e-16
```

Hasil model ketiga dengan variabel interaksi

Analysis of Variance Table

```
Model 1: Life.expectancy ~ Adult.Mortality + Diphtheria + status_Developed +
    HIV.AIDS + Polio + infant.deaths + Adult.Mortality:Diphtheria +
    Adult.Mortality:HIV.AIDS + Diphtheria:Polio
Model 2: Life.expectancy ~ Adult.Mortality + Diphtheria + Status_Developed +
    HIV.AIDS + Polio + infant.deaths + Adult.Mortality:HIV.AIDS +
    Adult.Mortality:Diphtheria + Diphtheria:Polio + Adult.Mortality:Polio
  Res.Df  RSS Df Sum of Sq  F    Pr(>F)
1     173 2605.3
2     172 2564.3   1    41.048 2.7533 0.09888 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hasil uji perbandingan model dengan atau tanpa variabel interaksi
Adult.Mortality:Polio

Model Akhir

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_1 x_4 + \beta_8 x_1 x_2 + \beta_9 x_2 x_5 + \beta_{10} x_1 x_5$$

Variabel:

$x_1 = \text{Adult.Mortality}$

$x_2 = \text{Diphtheria}$

$x_3 = \begin{cases} 1, & \text{Developed} \\ 0, & \text{Developing} \end{cases}$

$x_4 = \text{HIV.AIDS}$

$x_5 = \text{Polio}$

$x_6 = \text{infant.deaths}$

Interpretasi:

β_0 : Nilai intersep, ekspektasi y jika nilai setiap variabel prediktor adalah 0.

$\beta_1 + \beta_7 x_4 + \beta_8 x_2 + \beta_{10} x_5$: Kenaikan nilai ekspektasi y jika terjadi kenaikan satu unit Adult.Mortality, untuk nilai variabel prediktor lain termasuk x_2 : Diphtheria, x_4 : HIV.AIDS, dan x_5 : Polio yang konstan.

$\beta_2 + \beta_8 x_1 + \beta_9 x_5$: Kenaikan nilai ekspektasi y jika terjadi kenaikan satu unit Diphtheria, untuk nilai variabel prediktor lain termasuk x_1 : Adult.Mortality dan x_5 : Polio yang konstan.

β_3 : Kenaikan nilai ekspektasi y dari negara maju dibandingkan dengan negara berkembang, untuk nilai variabel prediktor lain yang konstan

$\beta_4 + \beta_7 x_1$: Kenaikan nilai ekspektasi y jika terjadi kenaikan satu unit HIV.AIDS, untuk nilai variabel prediktor lain termasuk x_1 : Adult.Mortality yang konstan.

$\beta_5 + \beta_9 x_2 + \beta_{10} x_1$: Kenaikan nilai ekspektasi y jika terjadi kenaikan satu unit Polio, untuk nilai variabel prediktor lain termasuk x_1 : Adult.Mortality dan x_2 : Diphtheria yang konstan.

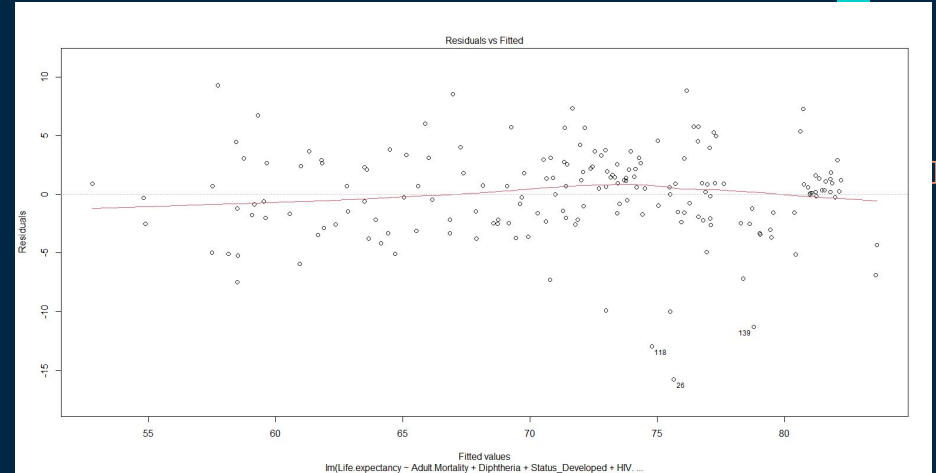
β_6 : Kenaikan nilai ekspektasi y jika terjadi kenaikan satu unit infant.deaths, untuk nilai variabel prediktor lain yang konstan.

Analisis Residual

Pengecekan asumsi heteroskedastisitas

Pengecekan keberadaan *outlier*

26	35	71	98	104
-4.196242	2.309487	-2.391986	-2.653360	-2.634448
118	126	139	144	183
-3.461779	2.647824	-2.980752	-2.040097	2.605349



Pengecekan asumsi normalitas

One-sample kolmogorov-smirnov test

```
data: rstandard(mod3)
D = 0.070173, p-value = 0.3284
alternative hypothesis: two-sided
```

Analisis Multikolinieritas

Adult.Mortality	Diphtheria	Status_Developed	HIV.AIDS	Polio
35.784681	21.268351	1.259682	34.185208	50.676567
infant.deaths	Adult.Mortality:HIV.AIDS	Adult.Mortality:Diphtheria	Diphtheria:Polio	Adult.Mortality:Polio
1.060557	41.444818	16.180462	46.443735	34.740036

Interesting findings from the model

Dapat dilihat bahwa outlier-outlier pada data kami adalah Burkina Faso, Chile, Haiti, Madagascar, Mauritania, Niger, Peru, São Tomé and Príncipe, Sierra Leone, dan Zimbabwe, di mana semua negaranya merupakan negara yang berkembang. Dapat dinyatakan bahwa kecenderungan terjadi kesenjangan pada negara-negara berkembang lebih tinggi dibandingkan negara-negara maju.

Kesimpulan

Kami menyimpulkan bahwa beberapa faktor yang mempengaruhi angka harapan hidup berdasarkan data yang kami miliki adalah tingkat mortalitas, imunisasi untuk diphtheria, tingkat perkembangan suatu negara, kematian akibat HIV/AIDS, imunisasi untuk polio, dan tingkat kematian bayi. Adapun faktor lain pada data kami seperti rata-rata pengeluaran penduduk untuk kesehatan dan kasus terjadinya campak tidak memiliki pengaruh yang besar dalam menentukan angka harapan hidup.

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, light blue, and light orange. Some of these squares are connected to the top edge of the frame by thin, vertical white lines, creating a sense of depth or falling objects. The central text is composed of two parts: 'TERIMA' in white and 'KASIH' in light orange.

TERIMA
KASIH